Karnataka State Akkamahadevi Women's University, Vijayapura

"Deepfake Video Analysis and Detection Using Artificial Intelligence"

Research Synopsis for the degree of

DOCTOR OF PHILOSOPHY

IN

DEPARTMENT OF COMPUTER SCIENCE

Submitted by

Mrs. Shilpa Kaman

(Reg. No: KSAWU/Ph.D/499/2020-21)

Under the Guidance of

Dr. Aziz Makandar

Professor

Department of Computer Science

KSAWU, Vijayapura

2021

Abstract

Fake multimedia has become a central problem in the last few years, especially after the advent Deepfakes. Deepfakes are realistic-looking fake media that has been generated by deep-learning algorithms which are being widely used to swap faces or objects in video and digital content. These images and videos are manipulated using advanced deep learning tools, like autoencoders (AE) or generative adversarial networks (GAN). Deepfake algorithms can create fake images and videos that humans cannot distinguish them from authentic ones. The boundary between real and synthetic media has become very thin. With this technology, creating realistic manipulated media assets may be very easy, provided one can access large amounts of data. Even though there are many applications of deepfakes including movie productions, photography, video-games, virtual reality and many more, the same technology can cause threats to privacy, democracy and national security. Deepfakes can be used for creating fake porn videos to blackmail people, or building fake-news campaigns to manipulate the public opinion, can also be used to falsify satellite images with non-existent landscape features for malicious purposes. In the long run, it may also reduce trust in journalism, media including serious and reliable sources.

Deepfake videos can be categorized into face-swap, lip-synching, face re-enactment/puppet-master, face synthesis and attribute manipulation. As the quality of deepfakes are increasing, there is a need to improve the performance of deepfake detection methods. The inspiration is that what AI has broken can be fixed by AI as well. Detection methods are still in their infancy, various methods have been proposed and evaluated, there are many limitations in the existing systems. More emphasis should be given to improve the quality of datasets during pre-processing, enhance binary classification to multiclass/multi-label scheme, temporal aggregation while building the detection model. The proposed study facilitates the development of new and more robust methods to deal with the increasingly challenging deepfakes using Artificial Intelligence.

Synopsis of the Research work submitted to the Department of Computer Science, KSAWU, Vijayapura.

1	Research Title	"Deepfake Video Analysis and Detection Using Artificial Intelligence"
2	Name of the Research Scholar	Mrs. Shilpa Kaman
3	Name and Designation of the Research Supervisor	Dr. Aziz Makandar Professor Department of Computer Science KSAWU, Vijayapura.
4	Ph.D Registration Number	KSAWU/Ph.D/499/2020-21
5	Date of Registration	23-09-2021

Signature of the Candidate

Signature of the Guide with seal

Signature of the Chairman with seal

Contents

Sl. No.	Topics	Page No.
1	Introduction	1
2	Literature Review	6
3	Problem Statement	9
4	Objectives	9
5	Methodology	10
6	Conclusion	11
	References	12

1. Introduction

The availability of economical digital smart devices like cell phones, tablets, laptops, and digital cameras has resulted in the exponential growth of multimedia content (e.g. images and videos) in cyberspace. Additionally, the evolution of social media over the last decade has allowed people to share captured multimedia content rapidly, leading to a significant increase in multimedia content generation and ease of access to it. At the same time, there is tremendous advancement in the field of Machine Leaning/Deep Learning with the introduction of sophisticated algorithms that can easily manipulate multimedia content to spread disinformation online through social media platforms. Given the ease with which false information may be created and spread, it has become increasingly difficult to know the truth and trust the information, which may result in harmful consequences. Disinformation is an active measure that has the potential to cause severe damage: election manipulation, creation of warmongering situations, defaming any person etc. [2].

Deepfakes are synthetic media in which a person in an existing image or video is replaced with someone else's likeness [3]. Deepfakes (stemming from “deep learning” and “fake”) are created by techniques that can superimpose face images of a target person onto a video of a source person to make a video of the target person doing or saying things the source person does [2]. In recent times, deepfake generation has significantly advanced and it could be used to propagate disinformation around the globe, and may pose a severe threat, in the form of fake news, in the future. Deepfakes are synthesized, AI-generated, videos and audio. The use of videos as evidence in every sector of litigation and criminal justice proceedings is currently the norm. A video admitted as evidence must be authentic and its integrity must be verified. This is expected to become a challenging task, especially as deepfake generation becomes more sophisticated. Catching the deepfake alarming problem, research community has focused on developing deepfake detection algorithms and numerous results have been reported [1]. Now let us look at the principles of deepfake algorithms and how deepfakes are created technically by deep learning approaches.

1.1 Deepfake Creation

Deep learning is well known for its capability of representing complex and high-dimensional data. Deepfakes are created technically by deep learning approaches such as autoencoder and generative adversarial network (GAN) [4][5][6] models.

Autoencoder: One variant of the deep networks with that capability is deep autoencoders, which have been widely applied for dimensionality reduction and image compression [33–35]. The first attempt of deepfake creation was FakeApp, developed by a Reddit user using autoencoder-decoder

pairing structure [36][37]. In that method, the autoencoder extracts latent features of face images and the decoder is used to reconstruct the face images. To swap faces between source images and target images, there is a need of two encoder-decoder pairs where each pair is used to train on an image set, and the encoder's parameters are shared between two network pairs. In other words, two pairs have the same encoder network. This strategy enables the common encoder to find and learn the similarity between two sets of face images, which are relatively unchallenging because faces normally have similar features such as eyes, nose, mouth positions. Fig. 2 shows a deepfake creation process where the feature set of face A is connected with the decoder B to reconstruct face B from the original face A.

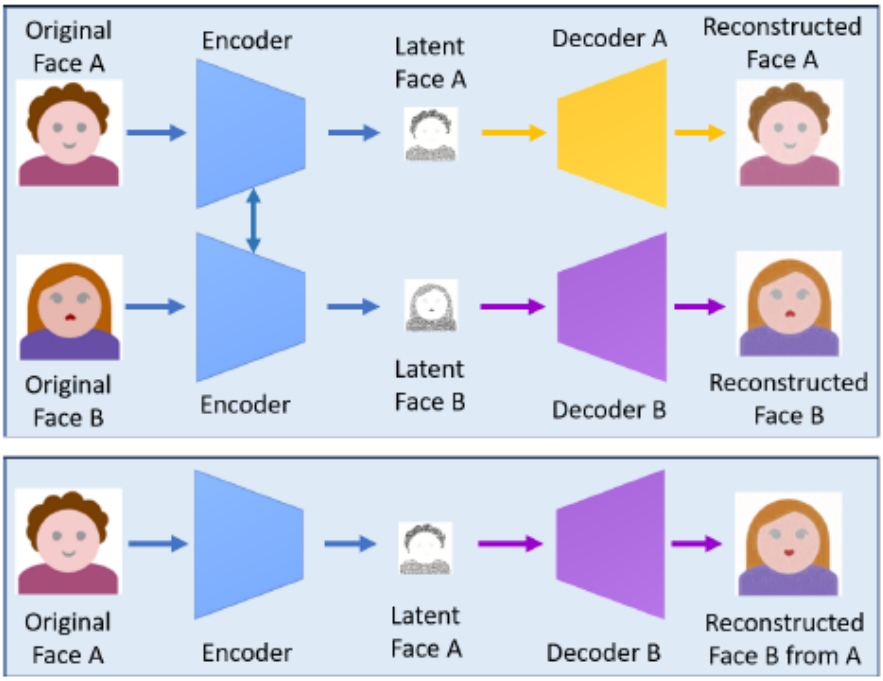

Fig. 1: A deepfake creation model using two encoder-decoder pairs [1].

Generative Adversarial Network (GAN): A conventional GAN model comprises two neural networks: a generator and a discriminator as depicted in Fig. 3. Given a dataset of real images x having a distribution of p_{data}, the aim of the generator G is to produce images G(z) similar to real images x with z being noise signals having a distribution of p_z. The aim of the discriminator G is to correctly classify images generated by G and real images x. The discriminator D is trained to improve its classification capability, i.e., to maximize D(x), which represents the probability that x is a real image rather than a fake image generated by G. After sufficient training, both networks improve their capabilities, i.e., the generator G is able to produce images that are really similar to real images while the discriminator D is highly capable of distinguishing fake images from real

ones. An overview of the popular deepfake creation tools is given in the following table 1.

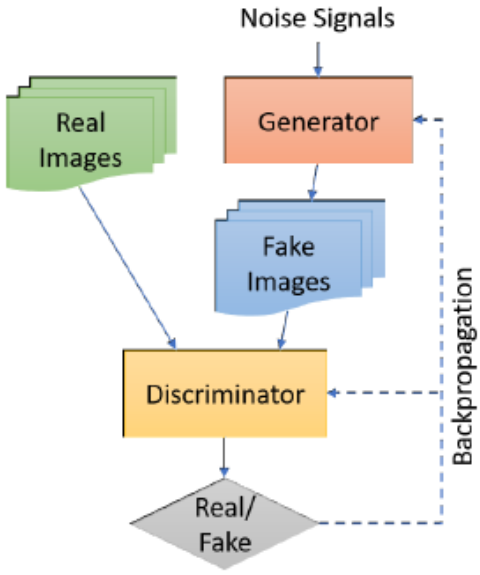

Fig. 2: The GAN architecture consisting of a generator and a discriminator, and each can be implemented by a neural network [1].

Table 1: Summary of Deepfake creation tools with their notable characteristics.

Sl. No.	Deepfake Creation Tools	Characteristics
1	Faceswap	-Using two encoder-decoder pairs. -Parameters of the encoder are shared [15].
2	Faceswap-GAN	Adversarial loss and perceptual loss (VGGface) are added to an auto-encoder architecture [17].
3	DiscoFaceGAN	-Generate face images of virtual people with independent latent variables of identity, expression, pose, and illumination. -Embed 3D priors into adversarial learning [19].
4	FSGAN	-A face swapping and reenactment model that can be applied to pairs of faces without requiring training on those faces. -Adjust to both pose and expression variations [20].
5	StyleGAN	-A new generator architecture for GANs is proposed based on style transfer literature. -The new architecture leads to automatic, unsupervised separation of high-level attributes and enables intuitive, scale-specific control of the synthesis of images [23].
6	DeepFaceLab	-Expand from the Faceswap method with new models, e.g. H64, H128, LIAEF128, SAE [21]. -Support multiple face extraction modes, e.g. S3FD, MTCNN, dlib, or manual [21].
7	DFaker	-DSSIM loss function [22] is used to reconstruct face. -Implemented based on Keras library.
8	DeepFake tf	Similar to DFaker but implemented based on tensorflow [18].
9	Face2Face	-Real-time facial reenactment of monocular target video sequence, e.g. Youtube video. -Animate the facial expressions of the target video by a source actor and re-render the manipulated output video in a photo-realistic fashion [24].

10	Neural Textures	-Feature maps that are learned as part of the scene capture process and stored as maps on top of 3D mesh proxies. -Can coherently re-render or manipulate existing video content in both static and dynamic environments at real-time rates [25].

1.2 Deepfake Detection

Deepfake detection methods are grouped into two major categories: fake image detection methods and fake video detection ones as shown in figure 1. The latter is distinguished into two smaller groups: visual artifacts within single video frame-based methods and temporal features across frames-based ones. Whilst most of the methods based on temporal features use deep learning recurrent classification models, the methods use visual artifacts within video frame can be implemented by either deep or shallow classifiers.

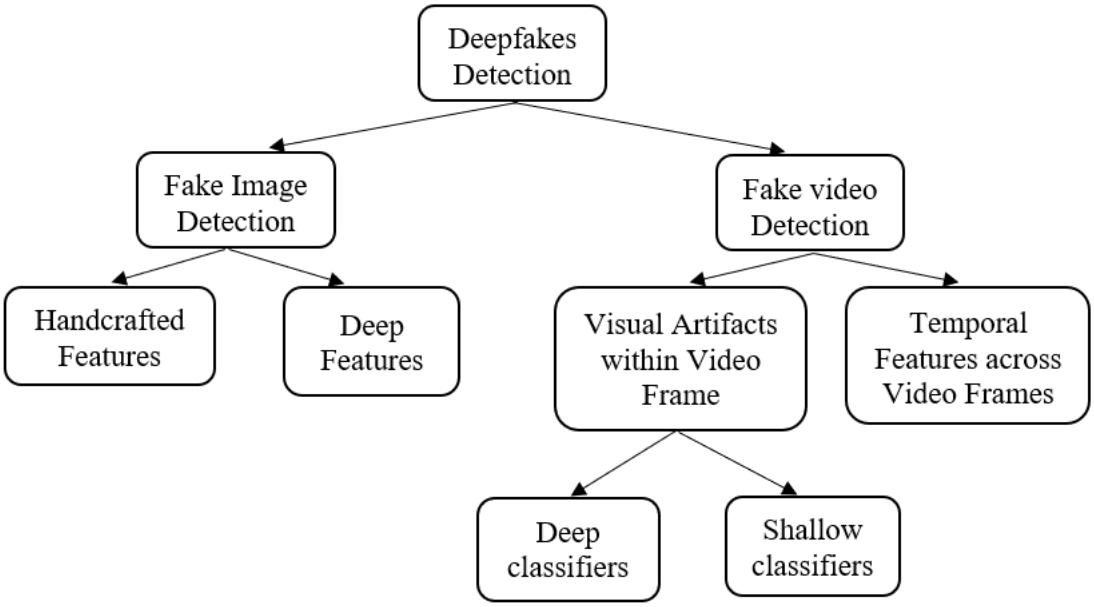

Fig. 3: Categories of Deepfake Detection methods.

1.2.1 Fake Image Detection

Digital images are particularly vulnerable to unauthorized access and illicit alterations. As a result, image credibility verification approaches are constantly gaining popularity due to their importance in a variety of social concerns, including the government, military, forensics, and electronic commerce, among others. Even a minor attack could have disastrous consequences. Therefore, protecting images from tampering has become increasingly crucial. Many approaches have been developed in the literature to ensure the authenticity and integrity of digital images. These methodologies can be grouped into active and passive approaches for image authentication and tampering detection. **Active image authentication** relies on prior information about the image under consideration, it has a general solution based on watermarking, signature and hybrid techniques.

Passive image authentication techniques are used when no prior information is available, they rely on the image's inherent natural characteristics to provide implicit authentication [7].

Early attempts were based on handcrafted features obtained from artifacts and inconsistencies of the fake image synthesis process. Recent methods, have commonly applied deep learning to automatically extract salient and discriminative features to detect deepfakes.

1.2.2 Fake Video Detection

Deepfakes videos can be categorized into the following types:

- **Face-swap:** In face-swap deepfakes, the face of the source person is replaced with the target person to generate a fake video of the target person, trying to portray actions to the target person which in reality the source person has done. Face-swap-oriented deepfakes are usually generated to target the popularity or reputation of famous personalities by showing them in scenarios in which they never appeared [8], and to damage reputations in the face of the public.
- **Lip-synching:** In lip-synching based deepfakes, the movements of the target person's lips are transformed to make them consistent with some specific audio recording. Lip-syncing is generated with the aim of showing an individual speaking in a way in which the attacker devises the victim to speak.
- **Face re-enactment/puppet-master:** With puppet-master, deepfakes are created by mimicking the expressions of the target person, such as eye-movement, facial expressions, and head movement. Puppet-master deepfakes aim to hijack the source person's expression, or even the full-body, [9] in a video, and to animate according to the impersonator's desire.
- **Face synthesis and attribute manipulation:** Face synthesis and attribute manipulation involves the generation of photo-realistic face images and facial attribute editing. This manipulation is generated to spread disinformation on social media using fake profiles.

Most of the image detection methods cannot be used for videos because of the strong degradation of the frame data after video compression. Furthermore, videos have temporal characteristics that are varied among sets of frames and they are thus challenging for methods designed to detect only still fake images.

- ***Temporal Features across Video Frames***

Spatiotemporal features of video streams are used to detect deepfakes. Video manipulation is carried out on a frame-by-frame basis so that low level artifacts produced by face manipulations are believed to further manifest themselves as temporal artifacts with inconsistencies across frames.

- ***Visual Artifacts within Video Frame***

This approach decomposes videos into frames and explores visual artifacts within single frames to obtain discriminant features. These features are then distributed into either a deep or shallow classifier to differentiate between fake and authentic videos.

2. Literature Review

The development of advanced deep neural networks and the availability of large amount of data have made the forged images and videos almost indistinguishable to humans and even to sophisticated computer algorithms. The process of creating those manipulated images and videos is also much simpler today as it needs as little as an identity photo or a short video of a target individual. Less and less effort is required to produce a stunningly convincing tempered footage. Recent advances can even create a deepfake with just a still image. Deepfakes therefore can be a threat affecting not only public figures but also ordinary people [10].

- Deepfakes were used to swap faces of celebrities or politicians to bodies in porn images and videos.
- It is threatening to world security when deepfake methods can be employed to create videos of world leaders with fake speeches for falsification purposes.
- Deepfakes therefore can be abused to cause political or religion tensions between countries, to fool public and affect results in election campaigns.
- Create chaos in financial markets by creating fake news.
- Deepfakes can be even be used to generate fake satellite images of the Earth to contain objects that do not really exist to confuse military analysts.
- Deepfakes can have creative or productive impacts in photography, video games, virtual reality, movie productions, and entertainment.

However, the number of malicious uses of deepfakes largely dominates that of the positive ones.

The earliest example of manipulated multimedia content occurred in 1860 when a portrait of southern politician John Calhoun was skilfully manipulated by replacing his head with that of US President Abraham Lincoln [11]. Usually, such manipulation is accomplished by adding (splicing), removing (inpainting), and replicating (copy-move) the objects within or between two images [12].

Then, suitable post-processing steps like scaling, rotating, and colour adjustment are applied to improve visual appearance, scale, and perspective coherence. The first true deepfake appeared online in September 2017 when a Reddit user named "deepfake" posted a series of computer-generated videos of famous actresses with their faces swapped onto pornographic content [13]. Another notorious deepfake case was the release of the deepNude application that allowed users to generate fake nude images [14].

Finding the truth in digital domain therefore has become increasingly critical. It is even more challenging when dealing with deepfakes as they are majorly used to serve malicious purposes and almost anyone can create deepfakes these days using existing deepfake tools. Detection methods are still in their early stage compared to deepfake creation techniques. Various methods have been proposed and evaluated for deepfake detection, some of these techniques are listed in the following table 2.1.

Table 2: Existing Deepfake video detection methods.

Sl. No.	Methods	Classifiers/ Techniques	Key Features	Dataset
1	MesoNet [26]	CNN	- Two deep networks, i.e. Meso-4 and MesoInception-4 are introduced to examine deepfake videos at the mesoscopic analysis level. - Accuracy obtained on deepfake and FaceForensics datasets are 98% and 95% respectively.	Two datasets: deepfake one constituted from online videos and the FaceForensics one created by the Face2Face approach [24].
2	Eye Blinking [27]	LRCN	- Use LRCN to learn the temporal patterns of eye blinking. - Based on the observation that blinking frequency of deepfakes is much smaller than normal.	Consist of 49 interview and presentation videos, and their corresponding generated deepfakes.
3	Eye, teach and facial texture [28]	Neural Network and Logistic Regression	Uses facial structure variations, as well as missing reflections and details in the eye and teeth areas of deepfakes. Classification is done using logistic regression and neural networks.	YouTube video dataset was downloaded to train model.
4	Using face Warping [29]	VGG16, ResNet models.	CNN models are used to find artifacts depending on resolution inconsistencies between the warping face region and side area.	UADFV dataset a collection of 49 actual and 49 faked videos totalling 32752 frames. Deepfake TIMIT.
5	Head poses [30]	SVM	68 landmarks in the facial region are used to extract features.	UADFV dataset, a collection of 49 actual and 49 faked videos totalling 32752 frames.

			Using the collected characteristics and classify them using SVM.	The DARPA MediFor GAN Image Video Challenge produced -241 actual images and 252 deep fake ones.
6	Intra-frame and temporal inconsistencies [31]	CNN and LSTM	CNN is used to collect frame-level features, which are then used to train the LSTM model to classify deepfake videos.	A combination of 600 videos taken from a variety of online sources.
7	Spatiotemporal features with RCN [32]	RCN	RCN, which combines the convolutional network DenseNet and the gated recurrent unit cells is used to investigate temporal differences across frames.	FaceForensics++ dataset consists of 1,000 videos.
8	Analysis of PRNU [33]	PRNU	Examination of noisy patterns caused by manufacturer flaws in light-sensitive sensors in digital cameras. Investigate the changes in PRNU pattern among real and deepfake videos, as face changing is thought to change regional PRNU patterns.	10 real and 16 deepfake videos were created by the writers with the help of DeepFaceLab.
9	Using appearance and behaviour [34]	Face and behavioural traits are used to define rules.	ResNet-101 is used to learn temporal, behavioural biometrics based on facial expressions and head movements, while VGG is used to obtain static face biometrics.	FaceForensics++, Google/Jigsaw deepfake detection dataset, DFDC and Celeb-DF are some of the datasets used by this model.
10	Spatiotemporal features with LSTM [35]	Convolutional bidirectional recurrent LSTM network	For extracting facial features, an XceptionNet CNN is utilized, while voice embeddings are created by stacking numerous convolution layers. Cross entropy and Kullback-Leibler divergence are the two loss functions employed here.	The ASVSpoof 2019 Logical Access audio dataset, as well as the FaceForensics++ and Celeb-DF (5,639 deepfake videos) datasets.
11	Emotion audiovisual affective cues [36]	Siamese network	Emotional and modal aspects Deepfake detection involves extracting embedding vectors for the face and audio.	DFDC and Deepfake TIMIT.
12	FakeCatcher [37]	CNN	Because biological signals are not well retained geographically and temporally in deepfakes, extract them from portrait films and utilize them as an implicit diagnostic of authenticity.	UADFV, FaceForensics, FaceForensics++, Celeb-DF and a new dataset of 142 videos, independent of the generative model, resolution, compression, content, and context.

3. Problem Statement

Although remarkable advancements have been made in the performance of deepfake detectors there are numerous concerns about current detection techniques that need attention.

- Quality of videos from the available datasets reveals several ambiguities in comparison to actual manipulated content found on the internet. These ambiguities are due to imperfect steps in the manipulation techniques.
- Binary classification scheme should be enhanced to multiclass/multi-label and local classification/detection at the frame level, to cope with the challenges of real-world scenarios.
- Lack of Explainability in Detection Methods due to their black-box nature.
- Existing deepfake detection methods do not consider temporal consistency between frames
- Most deepfake detection approaches are affected by social media laundering.

These are real challenges for detection method development. Hence the proposed system focuses on introducing more robust, scalable and generalizable deepfake video detection method.

4. Objectives

Main objectives of proposed research work are as follows

1) Identify the gap in the literature by studying different existing deepfake detection methods or algorithms and analyze their performance on predictions.
2) Collect the dataset comprising of both real and fake videos in right proportions.
3) Perform preprocessing of dataset to improve the quality by focusing on impact of preprocessing and aggregation of frame level to video level prediction.
4) Prepare the dataset to conduct experiments by applying augmentation, sampling and normalization techniques. Split the datasets into training and testing data.
5) Design and Develop Deepfake detection model to accomplish following tasks using Machine Learning and Deep Learning techniques.
    - Feature extraction to extract frame level features from videos.
    - Classification of real and fake videos.
6) Compare and analyze the results obtained from the proposed model with existing systems.
7) Assess and evaluate the model's performance using different parameters.

5. Methodology

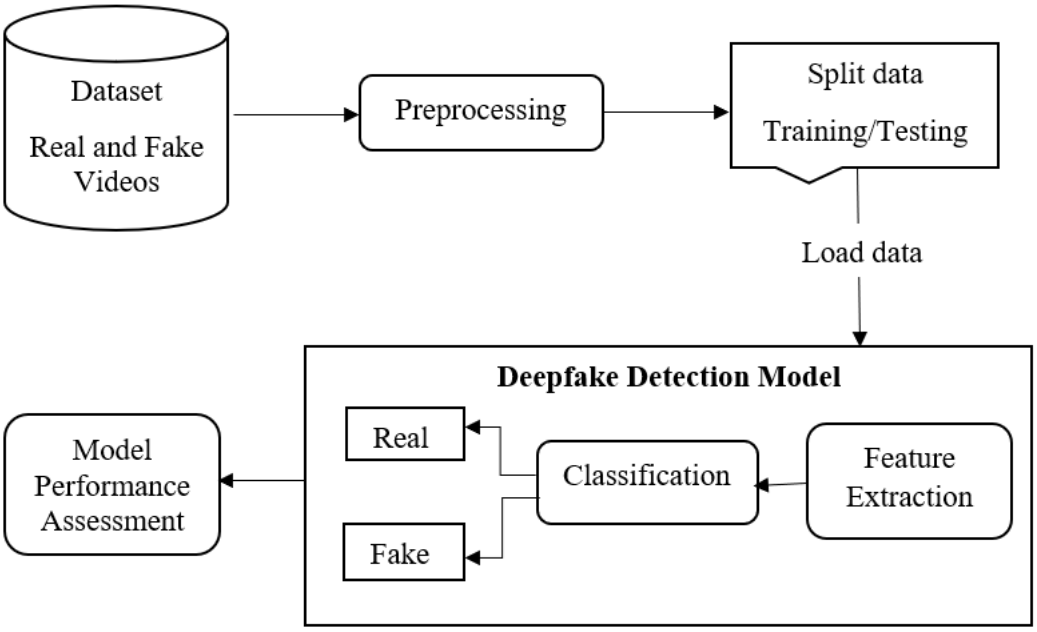

Fig. 4: Workflow of the proposed Deepfake Detection Model.

Existing Deepfake detection systems focus mostly on presenting their proposed architecture, and give less emphasis on data pre-processing and its impact on the final detection model. Hence preprocessing of data should be done carefully in order to remove ambiguities in the dataset thereby improve performance of the system. Preprocessing involves dividing the videos into frames, enhancing the quality of input frames, performing face detection and face cropping and preparing data for experiments. Preprocessed data is split into training and testing samples and loaded to model. Design and develop deepfake detection model which accomplish the tasks of feature extraction and classification with the help of machine learning and deep leaning algorithms. Evaluate the performance of the model using different parameters.

6. Conclusion

Deepfakes open new possibilities in digital media. They can be used for creation of hyper-realistic Computer-Generated Imagery (CGI), Virtual Reality (VR), Augmented Reality (AR), Education, Animation, Arts, Cinema, Robotics and many other fields. On another spectrum, they are technologies that can cause havoc and distrust to the general public. Since deepfakes are deceptive in nature, they can be used for malicious purposes. Over the years, significant research has been devoted in developing new techniques for detecting different types of deepfakes, they all suffer from their share of limitations. Some of the challenges identified in the current deepfake detection approaches are (i) Quality of deepfake datasets (ii) Performance evaluation (iii) Temporal aggregation (iv) Lack of explainability in detection models (v) Social media laundering. By considering these drawbacks of existing methods the proposed system aims at developing more robust, reliable and generalizable deepfake detection model.

References

[1] Rahman, Ashifur, et al. "A Qualitative Survey on Deep Learning Based Deepfake Video Creation and Detection Method." Aust. J. Eng. Innov. Technol 4.1 (2022): 13-26.

[2] Masood, Momina & Nawaz, Marriam & Malik, Khalid & Javed, Ali & Irtaza, Aun. (2021). Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward.

[3] https://en.wikipedia.org/wiki/Deepfake

[4] Ming-Yu Liu, Xun Huang, Jiahui Yu, Ting-Chun Wang, and Arun Mallya. Generative adversarial networks for image and video synthesis: Algorithms and applications. Proceedings of the IEEE, 109(5):839–862, 2021.

[5] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. arXiv preprint arXiv:1511.05644, 2015.

[6] Ayush Tewari, Michael Zollhoefer, Florian Bernard, Pablo Garrido, Hyeongwoo Kim, Patrick Perez, and Christian Theobalt. High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42 (2):357–370, 2018.

[7] Chennamma, Hr & Madhushree, B. (2022). A comprehensive survey on image authentication for tamper detection with localization. Multimedia Tools and Applications. 10.1007/s11042-022-13312-1.

[8] J. F. Boylan, "Will deep-fake technology destroy democracy?" The New York Times, Oct, vol. 17, 2018.

[9] C. Chan, S. Ginosar, T. Zhou, and A. A. Efros, "Everybody Dance Now," in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 5933-5942.

[10] S. P and S. Sk, "DeepFake Creation and Detection:A Survey," *2021* Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 584-588, doi: 10.1109/ICIRCA51532.2021.9544522.

[11] N. A. Mhiripiri and T. Chari, Media law, ethics, and policy in the digital age. IGI Global, 2017.

[12] L. Verdoliva, "Media forensics and deepfakes: an overview," arXiv preprint arXiv:2001.06564, 2020.

[13] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep Learning for Deepfakes Creation and Detection," arXiv preprint arXiv:1909.11573, 2019.

[14] J. Vincent. (18.09.2020). New AI deepfake app creates nude images of women. Available:

https://www.theverge.com/2019/6/27/18760896/deepfake-nude-ai-app-women-deepnudenon-consensual-pornography.

[15] Faceswap: Deepfakes software for all. https://github.com/deepfakes/faceswap.

[16] FakeApp 2.2.0. https://www.malavida.com/en/soft/fakeapp/.

[17] Faceswap-GAN. https://github.com/shaoanlu/faceswap-GAN.

[18] DeepFake tf: Deepfake based on tensorflow. https://github.com/StromWine/DeepFake_tf.

[19] Yu Deng, Jiaolong Yang, Dong Chen, Fang Wen, and Xin Tong. Disentangled and controllable face image generation via 3D imitative-contrastive learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5154–5163, 2020.

[20] Yuval Nirkin, Yosi Keller, and Tal Hassner. FSGAN: Subject agnostic face swapping and reenactment. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 7184–7193, 2019.

[21] DeepFaceLab: Explained and usage tutorial. https://mrdeepfakes.com/forums/thread-deepfacelab-explained-and-usage-tutorial.

[22] DSSIM. https://github.com/keras-team/keras-contrib/blob/master/keras _contrib/losses/dssim.py.

[23] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4401–4410, 2019.

[24] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2Face: Real-time face capture and reenactment of RGB videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2387–2395, 2016.

[25] Justus Thies, Michael Zollh¨ofer, and Matthias Nießner. Deferred neural rendering: Image synthesis using neural textures. ACM Transactions on Graphics (TOG), 38(4):1–12, 2019.

[26] Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. MesoNet: a compact facial video forgery detection network. In 2018 IEEE InternationalWorkshop on Information Forensics and Security (WIFS), pages 1–7. IEEE, 2018.

[27] Yuezun Li, Ming-Ching Chang, and Siwei Lyu. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS), pages 1–7. IEEE, 2018.

[28] Falko Matern, Christian Riess, and Marc Stamminger. Exploiting visual artifacts to expose deepfakes and face manipulations. In IEEE Winter Applications of Computer Vision

Workshops (WACVW), pages 83–92. IEEE, 2019.

[29] Yuezun Li and Siwei Lyu. Exposing deepfake videos by detecting face warping artifacts. arXiv preprint arXiv:1811.00656, 2018.

[30] Xin Yang, Yuezun Li, and Siwei Lyu. Exposing deep fakes using inconsistent head poses. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 8261–8265. IEEE, 2019.

[31] David G¨uera and Edward J Delp. Deepfake video detection using recurrent neural networks. In 15th IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS), pages 1–6. IEEE, 2018.

[32] Ekraam Sabir, Jiaxin Cheng, Ayush Jaiswal, Wael AbdAlmageed, Iacopo Masi, and Prem Natarajan. Recurrent convolutional strategies for face manipulation detection in videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 3(1):80–87, 2019.

[33] Marissa Koopman, Andrea Macarulla Rodriguez, and Zeno Geradts. Detection of deepfake video manipulation. In The 20th Irish Machine Vision and Image Processing Conference (IMVIP), pages 133–136, 2018.

[34] Shruti Agarwal, Hany Farid, Tarek El-Gaaly, and Ser-Nam Lim. Detecting deep-fake videos from appearance and behavior. In IEEE International Workshop on Information Forensics and Security (WIFS), pages 1–6. IEEE, 2020.

[35] Akash Chintha, Bao Thai, Saniat Javid Sohrawardi, Kartavya Bhatt, Andrea Hickerson, Matthew Wright, and Raymond Ptucha. Recurrent convolutional structures for audio spoof and video deepfake detection. IEEE Journal of Selected Topics in Signal Processing, 14(5):1024–1037, 2020.

[36] Trisha Mittal, Uttaran Bhattacharya, Rohan Chandra, Aniket Bera, and Dinesh Manocha. Emotions don’t lie: A deepfake detection method using audio-visual a_ective cues. arXiv preprint arXiv:2003.06711, 3, 2020.

[37] Umur Aybars Ciftci, Ilke Demir, and Lijun Yin. FakeCatcher: Detection of synthetic portrait videos using biological signals. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.